Article

Not peer-reviewed version

Comparison of Health and Data Science Methodologies for Identifying Avoidable Hospitalizations in Patients with Diabetes: An Interdisciplinary Approach

Carlos Hernandez-Nava * , Miguel-Felix Mata-Rivera , Sergio Flores-Hernandez

Posted Date: 28 May 2024

doi: 10.20944/preprints202405.1844.v1

Keywords: Methodology; Data Science; Public health; standardization; epidemiology, datasets, hospital discharges, adjusted rate

Article

Comparison of Health and Data Science Methodologies for Identifying Avoidable Hospitalizations in Patients with Diabetes: An Interdisciplinary Approach

Carlos Hernández-Nava [1]*, Miguel-Félix Mata-Rivera [1] and Sergio Flores-Hernández[2]

[1] Interdisciplinary Professional Unit in Engineering And Advanced Technologies of the National Polytechnic Institute, 07340 Gustavo A Madero, Mexico City, Mexico; hernandeznc@ipn.mx (C.H.); mmatar@ipn.mx (M.M.)

[2] National Public Health Institute of Mexico, 62100 Cuernavaca, Morelos, Mexico; sergio.flores@insp.mx (S.F.)

* Correspondence: hernandeznc@ipn.mx

Abstract: In the whole world, including Mexico, the rising prevalence of diabetes poses significant challenges to healthcare systems, with a notable impact on hospital admissions even being an ambulatory care sensitive condition, that means, hospital admissions are avoidable. Traditional healthcare methodologies have been instrumental in managing diabetes and preventing complications, yet they often face limitations such as merging, cleansing, and outlier analysis of health data to identify and address diseases effectively. This paper aims to address this gap by conducting a comprehensive comparison of the methodologies used in healthcare and data science for this purpose. This work uses hospital diabetes discharge records from 2010 to 2023, a total of 36,665,793 records, which belong to medical units of the Ministry of Health of Mexico. In this work, we seek to provide the arguments why as a data scientist it is mandatory to learn the field of knowledge of the problem and its implications if this is not done, and therefore to disclose insights that can help in policy decisions and reduce the burden of avoidable hospitalizations. The approach is mainly based on the standardization or adjust rates by sex and age groups. This study provides the foundations for a new way of data scientist must deal with health data.

Keywords: methodology; data science; public health; standardization; epidemiology; datasets; hospital discharges; adjusted rate

1. Introduction

According to INEGI, Mexico is among the leading countries worldwide in the prevalence of diabetes, with approximately 10.3 million adults diagnosed with the condition [1]. Furthermore, the burden of avoidable hospitalizations related to complications of diabetes mellitus (AHRDM) remains a critical issue, imposing substantial economic costs and straining healthcare resources; in 2016, the cost was 1,563 million dollars [2].

Noncommunicable diseases in the world kill each year 74% of all deaths, according to the World Health Organization (WHO) [5]. Diabetes Mellitus (DM) is an ambulatory care sensitive condition (ACSC), which means that if the patient is treated effective and timely the hospitalization can be avoided. This is related to the quality of ambulatory care (also called primary care). That is the reason why this kind of hospitalization is called potentially avoidable.

Traditional healthcare methodologies have been instrumental in managing diabetes and preventing complications [3][4], but they often face limitations in identifying and addressing AHRDM effectively. In recent years, the emergence of data science methodologies has offered new avenues for understanding disease patterns and optimize healthcare delivery. Leveraging advanced analytics, and big data science presents a promising approach to complement traditional healthcare methods and improve patient outcomes.

Despite these advancements, there remains a gap in the literature regarding the comparative effectiveness of health and data science methodologies in identifying and mitigating AHRDM in patients, particularly within the context of the mexican healthcare system. Our study provides the basis for a new way to combine methodologies used in healthcare and data science. Through this interdisciplinary analysis, we seek to provide insights that can help make policy decisions and reduce the burden of AHRDM in Mexico.

According to the 10th IDF Diabetes Atlas issued by the International Diabetes Federation in 2021, there were 14,123,200 cases of diabetes in adults in Mexico [6], ranked eighth in the world, and occupied the same place in diabetes-related health expenditures, with 19.9 billion dollars in adults (20-79 years).

Statistics and Geography National Institute of Mexico, in the press release pertaining to World Day of Diabetes [1], presents the following numbers: 13% of deaths were due to diabetes, 51 % of men (71,330) and 49% of women (69,396), the mortality rate related to diabetes was 11 per 10,000 habs. Furthermore, 10.3% of the national population of 20 years and older were diagnosed with diabetes. The state of Mexico was ranked fourth and Mexico City ninth in the mortality rate by diabetes, all these numbers for the year 2021.

In fact, there exists a health indicator specified by the Agency for Healthcare Research and Quality (AHRQ) [7], the prevention quality indicator (PQI) number 93 which is a composite indicator, this indicator is a rate, where the numerator is hospital discharges with a principal diagnosis code ICD-10-CM for any of the following cases: diabetes with short-term complications (PQI #1), diabetes with long-term complications (PQI #3), uncontrolled diabetes (PQI #14) or lower extremity amputation among patients with diabetes (PQI #16).

All datasets records that match one of the PQI #93 specification codes with the principal diagnosis of the registered variable are avoidable hospitalizations related to diabetes mellitus (AHRDM) mentioned in the following sections. The denominator is the study population, in this work, people 20 years and older from Mexico City and the State of Mexico, this geographical area is called Metropolitan Area (MA, for short) for this article, one of the crowded areas in the world, in 2023 the study population was 19,219,280.

The results of this work will have serious implications; In particular, in Mexico City and the State of Mexico, the AHRDM rates decrease from 47.21 in 2010 to 39.05 in 2019, which means that 8 women in 2019 were treated effectively and timely manner compared to 2010 and avoid an hospitalization due to diabetes mellitus; in the same period for men only were 2; that is, before January 2020, the beginning of the COVID-19 pandemic period and from that month up to December 2023, for women, the AHRDM rates were 21.23 in 2020 and 30.12 in 2023; for men 32.38 in 2020 and 31.42 in 2023 (all rates per 100,000 habs.), which means that in four years (January 2020-December 2023) 10 women could not avoid reaching the hospital; for men, the rate do not show a significant change.

AHRDM rates are important because these are prevention quality indicators, that is, one way to assess the primary healthcare system, those rates can be used by professionals to make better decisions about healthcare systems.

In the epidemiology field, a design study must be defined; since the dataset already was created the study is called a retrospective analysis through a secondary analysis of databases, is called retrospective because the study uses past records and secondary because the databases were not made to this goal [9][8]. Records are hospital discharges for adults 20 years and older with DM between 2010 and 2023.

The main aim of the work is to provide the arguments why data scientists should be interdisciplinaries; otherwise, the results obtained by them are going to be useless. In addition, this research will benefit the data scientist as a guide for the analysis of health data. As a collateral goal, this work contributes to the field of data science by remarking the differences between the data science and public health methodologies to solve problems.

2. Methods

This section presents the methodology comparison, data sources, the process of data cleansing and preprocessing, outliers analysis, specific and crude rates, and direct standardization. Crude rates are those rates without an standardization and are the rates computed by data scientist in a naive manner, on the other hand, adjusted-age rates by sex are used in the "health world". Methodologies for both rates are presented in this section.

2.1. Methodology

The general known data science methodology involves the stages: data collection, preprocessing, modeling, interpretation; and presentation [10]. Where the common objective is to reach conclusions that help support decisions through techniques, outlier analysis, preprocessing data, and using algorithms for modeling data, among others.

But for public health problems, where the objective of studies is to provoke social and health impact, it is mandatory to follow the guidelines established by physicians, such as epidemiologists. Some of the differences between the implementation of data science in another problem and in public health are shown in Figure 1.

Mainly, the data standardization [12] is essential because it is necessary to adjust or transform some variables, but in public health it is not only that the values of some variables are within a range, the scaling process; or that the variables have the same mean and standard deviation, the normalization process. In particular, in public health it is crucial to standardize [13] the data because when comparing the results and concluding about the experiments, the values can be used by other scientists, not depending on the geographical area, country, or study population characteristics [14][17].

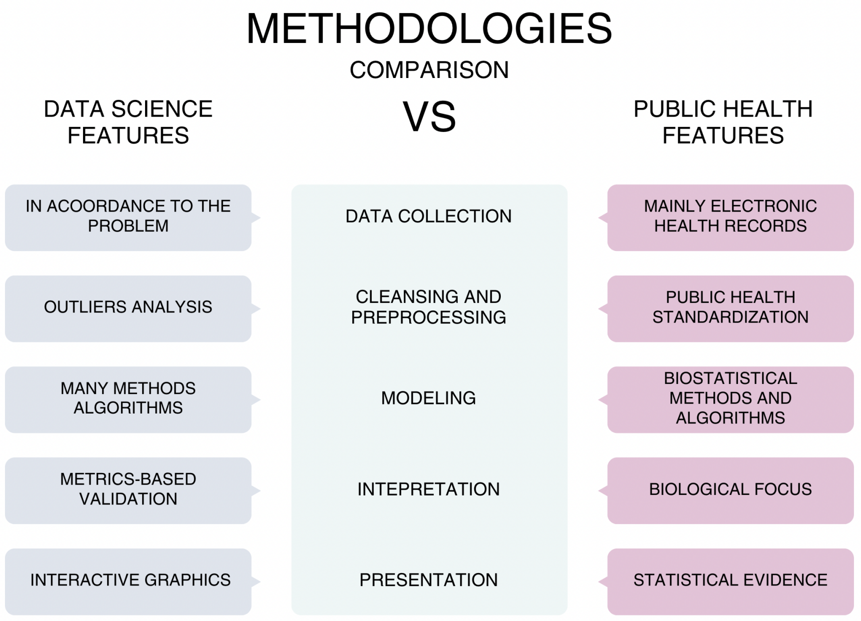

Figure 1. Data Science vs Public Health Methodologies Comparison

Figure 1 shows in the center each phase of a general, the main differences are the following:

1. **Data Collection**. Electronic Health Records are the basis of the data, but it is not limited, because some surveys, interviews could be included;
2. **Cleansing and preprocessing**, the main difference is the standardization in public health, which stratifies data and adjust directly or indirectly, in both methodologies manage values missings, data imputation, outliers analysis are applied;

3. **Modeling**, mainly in public health only use biostatistical methods and algorithms of regression;
4. **Interpretation**, in a generic data science project usually metrics are used while in a piblic health project interpretation must have a biological focus;
5. **Presentation**, this phase is where most mistakes are made by data scientists when deal with health data, because usually an interactive and friendly graphics are used but health professionals need statistical evidence like confidence interval coming from statisitical test or biostatistical methods.

2.2. Data Sources

The main data source is the open dataset provided by The Ministry of Health of Mexico, through the General Department of Health Information (SS / DGIS), which offers open data sets of deaths, hospital discharges, births, maternal deaths, and emergencies. In this work, the study utilize hospital discharge records from 2010 to 2023, a total of 36,665,793 records, which belong to medical units of the Ministry of Health of Mexico. Table 1 shows the national records by year, the records after the cleaning process, the corresponding records of the MA, then filters only the AHRDM cases and finally the population of 20 years and older. From the third column, each percentage corresponds to the value of the previous column.

Table 1. Hospital discharges records from Metropolitan Area (MA).

Year	National records	After cleaning	MA records	MA diabetes cases	Population habs.*
2010	2,634,339	2,632,251(99.92%)	510,888(19.41%)	7,071(1.38%)	15,160,396
2011	2,775,189	2,774,330(99.97%)	536,111(19.32%)	8,089(1.51%)	15,472,618
2012	2,880,706	2,880,075(99.98%)	566,765(19.68%)	7,455(1.32%)	15,784,839
2013	2,879,313	2,879,052(99.99%)	576,107(20.01%)	7,109(1.23%)	16,097,061
2014	2,959,197	2,958,924(99.99%)	603,358(20.39%)	7,074(1.17%)	16,409,284
2015	2,970,812	2,970,483(99.99%)	581,673(19.58%)	6,675(1.15%)	16,721,507
2016	2,955,144	2,952,697(99.92%)	599,528(20.30%)	7,478(1.25%)	17,033,727
2017	2,729,341	2,715,873(99.51%)	535,983(19.74%)	7,538(1.41%)	17,345,950
2018	2,623,379	2,622,560(99.97%)	530,078(20.21%)	7,530(1.42%)	17,658,172
2019	2,629,434	2,628,771(99.97%)	515,396(19.61%)	8,329(1.62%)	17,970,393
2020	1,937,344	1,934,458(99.85%)	387,942(20.05%)	4,767(1.23%)	18,282,615
2021	2,088,780	2,088,352(99.98%)	391,540(18.75%)	4,371(1.12%)	18,594,837
2022	2,203,636	2,197,685(99.73%)	372,418(16.95%)	5,643(1.52%)	18,907,058
2023	2,399,179	2,390,869(99.65%)	381,771(15.97%)	6,011(1.57%)	19,219,280
Total	**36,665,793**	**36,626,380**(99.89%)	**7,089,558**(19.36%)	**95,140**(1.34%)	**240,657,737**

*People of 20 years and older.

2.3. Data Cleansing and Preprocessing

In order to make this study reproducible, it is mentioned that the language used to clean and preprocess data is R, with R Studio as the IDE, to load data depends on the year of the file, even if all datasets are CSV files, each file has its own characteristics, which is the reason why the delimiter symbol and the number of variables (columns of the tables) differ from each other. For example, the separator symbol is the comma for the 2010 file year, giving the following code sentence:

```
read.csv(file = "data/open/EGRESO_2010.csv", header = TRUE, sep = ’,’)
```

In the same example, the number of variables is: `11,13,17,18,39` which correspond to the variables `EDAD`, `SEXO`, `ENTIDAD`, `MUNIC` and `AFECPRIN` consecutively. `EDAD` is the patient age, `SEXO` is the patient sex, 1 for men and 2 for women, `ENTIDAD` is the state name of the patient, `MUNIC` is the municipality name where the patient resides, and `AFCPRIN` contains the ICD-10-CM registered at hospital discharge.

2.4. Outliers Analysis

An outliers analysis is implemented through a hypothesis testing method, and some records with patients 999 years old and others with values different from 1 or 2 in the variable sex, being the allowed values, 1 for men and 2 for women. This analysis is crucial because databases usually have missing, redundant, or spurious values that could generate wrong outcomes.

The cleaning process includes a filtering process using regular expressions of the correct ICD-10-CM codes. Taking only the first four alphanumeric values with the regular expression in the sentence `grepl("[A-Z][0-9][0-9][A-Z0-9]", AFECPRIN)`.

2.5. Crude Rate

As a data scientist with naive knowledge of public health, the AHRDM rate is equal to the division of all cases (as numerator) by the study population of 20 years and older (as denominator); this is called the crude rate. Figure 2 shows the process that begins to collect discharge hospitalization records from SS/DGIS as a source, followed by a cleaning process and filtering of matching cases between MA discharge hospitalizations and the PQI #93 ICD-10-MD code. With these counted cases as numerator, divide by the MA population of 20 years and older to get the crude rate.

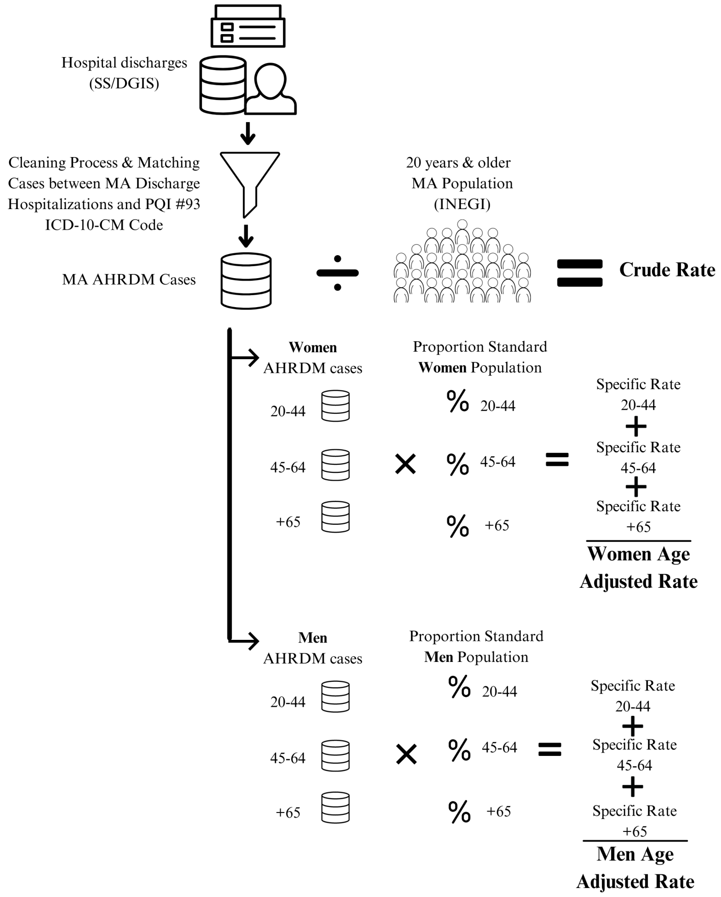

Figure 2. Process to Compute Crude Rate and Age Adjusted Rates by Sex.

2.6. Direct Standardization

Crude rates could be misleading, since the crude data do not take some characteristics of the population, the most common variables are the sex and age of the MA population. Figure 2 shows how to calculate the crude rate, from top to bottom, then divide by sex and age groups to calculate age-specific rates for women and men, finally getting the age adjusted rates using the standard population. To standardize, the direct method uses a population with a common age structure as a standard population, which in this work is the country population. The age groups are 20-44 years (young adults), 45-64 years (adults), and 65 years and older (elder adults). These age groups are divided into men and women.

The main purpose of standardization is to compare rates across time and geographical areas; for example, in the European Union, the proportion of people 65 years and older increased 5 percentage points, from 16% to 21% during the period of 2002-2022 [11], people of the Member States are getting older. Thus, to be fair with the comparison between the rates with the study population of this work, which is a proportion of people 65 years and older, 10%, half of the European Union.

2.7. Specific Rates

Specific rates are important because they are an absolute measurement and also useful for comparison and can be calculated for a specific group of people. Table 2 shows the sex-specific rates of AHRDM for the study population of this work. First of all, per year, the cases are grouped by women and men, also by the population, then to get the rates, divide the cases by population.

Table 2. Sex Specific Rates of AHRDM.

Year	Cases		Population (habs.)		Crude Rates (x100,000 habs.)	
	Women	Men	Women	Men	Women	Men
	(A)	(B)	(C)	(D)	(A/C)	(B/D)
2010	3,611	3,460	8,010,757	7,149,639	45.08	48.39
2011	4,001	4,088	8,173,368	7,299,250	48.95	56.01
2012	3,360	3,795	8,335,979	7,448,860	43.91	50.95
2013	3,544	3,565	8,498,589	7,598,472	41.70	46.92
2014	3,354	3,720	8,661,201	7,748,083	38.72	48.01
2015	3,276	3,399	8,823,812	7,897,695	37.13	43.04
2016	3,565	3,913	8,986,422	8,047,305	39.67	48.63
2017	3,544	3,994	9,149,034	8,196,916	38.74	48.73
2018	3,678	3,852	9,311,644	8,346,528	39.50	46.15
2019	3,943	4,386	9,474,255	8,496,138	41.62	51.62
2020	2,173	2,594	9,636,866	8,645,749	22.55	30.00
2021	2,023	2,348	9,799,477	8,795,360	20.64	26.70
2022	2,607	3,036	9,962,088	8,944,970	26.17	33.94
2023	3,314	2,697	10,124,698	9,094,582	32.73	29.65

3. Results

To obtain the AHRDM rates, the first step is to clean outliers, impute data if necessary, or delete records; then only the cases of MA were selected, the percentage range is from 1.12% in 2021 to 1.62% in 2019. With all AHRDM cases counted (numerator) and the study population per year (denominator), it is possible to obtain the crude rate (per 100,000 population). Cases and populations of MA diabetes per year are shown in Table 1. Crude rate is calculated by dividing the AHRDM cases by the study population; for example, in 2014 the cases were 7,074 (numerator) and the study population was 16,409,284 habs. (denominator); therefore, the crude rate for that year is $\frac{7{,}074}{16{,}409{,}284} = 43.11$ per $100,000$ habs.. In this way, the crude rates are plotted in Figure 3.

Figure 3 shows the crude rate (pink), where it is clearly possible to perceive the effects of COVID-19 from 2020 to now; in 2020 the rate was 26.07 and 23.51 for 2022, almost half of the crude rate of 2019, 46.35 AHRDM per 100,000 hab. Another thing to highlight is that before the pandemic period the average was 44.97 and from 2020 to 2023 was 27.68. But this rate is increasing slowly as the years go by.

Sex-specific rates of AHRDM are computed by the specific study population, dividing the cases by men and women. For example, in 2020, there were 2,173 cases of AHRDM of women and 2,594 cases of men, if these values are divided by the specific study population of women and men, the specific rates for women is 22.55 AHRDM and for men 30 AHRDM, both per 100,000 habs. (See Table 2 for each year).

Figure 3 also shows the sex-specific rates (red for men and blue for women); these are also known as adjusted rates by sex. It is possible to see that the rates for men are higher, except for 2023. On average before the pandemic period, the difference between AHRDM rates among men and women was 9.

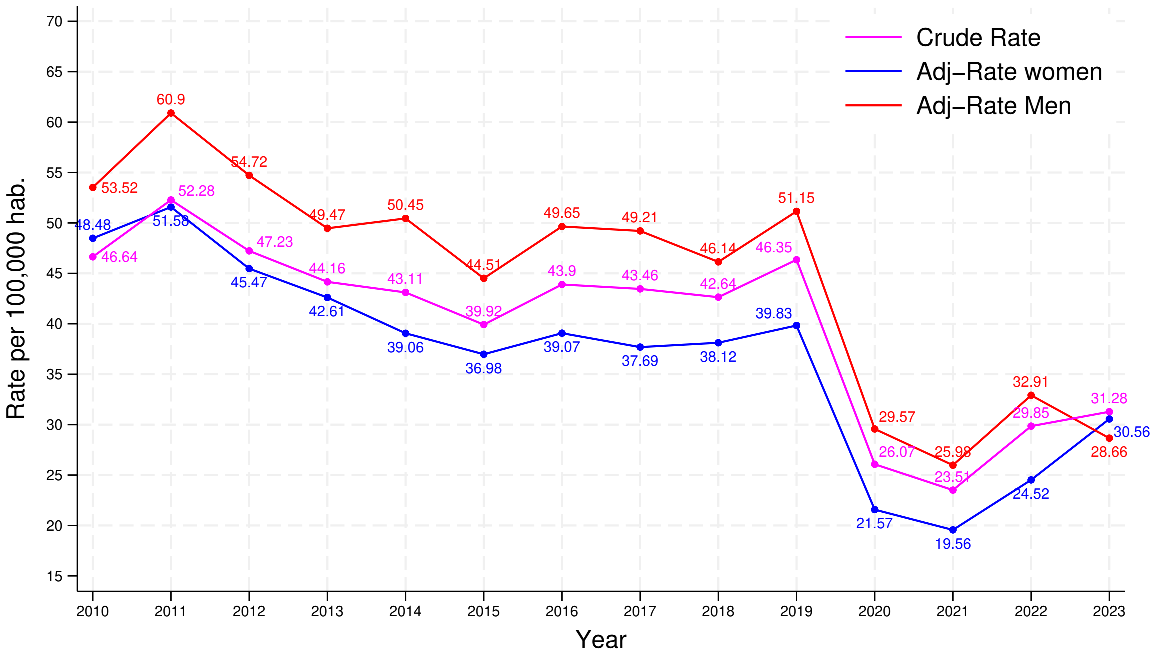

Figure 3. Crude Rate and Age Adjusted Rates by Sex.

In the process to compute rates, the data were adjusted by sex, but population also has biological characteristics by age group, for that reason, to customize the analysis, records are divided by age group, 20-44, 45-64 and 65 years and older. Table A1 shows the diabetes cases and populations per age group.

The literature related to stratification suggests that to divide age groups and sex to adjust data, this process is also known as stardardization, it is possible to get the crude rate by each group, second column in the Table 3, but to compute the specific rate, it is required to multiply the crude rate by the proportion of the standard population. That step is where the data are actually adjusted or standardized, even though this process is not hard, which is crucial from the point of view of public health.

A new way to deal with health databases involved five pillars, the first, the ability to manage diverse data sources and formats (CSV the most commmon); the second, knowledge of the variables and the foundations of the field where data science is applied; the third, the foundations of statistics and programming; the fourth, superb skills to clean, imputate, analyze, and all related to process data; at the final, the interpretation of the results, this will only be valuable when data scientist involved with the field of application.

Table 3. Age Adjusted Rates of AHRDM.

Year	Crude Rate		Specific Rate*		Age & Sex Adjusted	
Years-old	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**
2010						
20-44	15.49	19.76	9.17	12.27		
45-64	80.47	93.85	24.01	31.03	47.21	58.76
+65	127.97	115.83	14.03	15.45		
2011						
20-44	17.44	22.46	10.32	13.95		
45-64	88.86	109.03	26.52	36.10	50.37	66.86
+65	123.36	126.03	13.53	16.81		
2012						
20-44	16.65	19.69	9.86	12.23		
45-64	77.56	98.51	23.15	32.58	44.47	60.08
+65	104.47	114.51	11.46	15.27		
2013						
20-44	15.60	19.02	9.23	11.81		
45-64	70.68	90.40	21.09	29.90	41.62	54.30
+65	103.03	94.36	11.30	12.59		
2014						
20-44	13.61	18.16	8.05	11.28		
45-64	65.39	86.89	19.51	28.74	38.11	55.41
+65	96.23	115.40	10.55	15.39		
2015						
20-44	14.11	16.47	8.35	10.23		
45-64	58.86	79.32	17.56	26.23	36.09	48.87
+65	92.81	93.02	10.18	12.41		
2016						
20-44	16.51	19.29	9.77	11.98		
45-64	63.28	88.56	18.88	29.29	38.25	54.49
+65	87.57	99.12	9.60		13.22	
2017						
20-44	14.62	18.56	8.65	11.52		
45-64	62.90	89.86	18.77	29.72	36.87	54.01
+65	86.18	95.76	9.45	12.77		
2018						
20-44	16.28	17.79	9.64	11.04		
45-64	62.81	84.20	18.74	27.85	37.37	50.63
+65	81.96	88.02	8.99	11.74		
2019						
20-44	17.52	21.61	10.37	13.42		
45-64	64.76	91.32	19.32	30.20	39.05	56.09
+65	85.34	93.48	9.36	12.47		
2020						
20-44	11.44	14.85	6.77	9.22		
45-64	33.31	49.13	9.94	16.25	21.23	32.38
+65	41.19	51.77	4.52	6.91		
2021						
20-44	10.02	10.88	5.93	6.76		
45-64	29.93	48.01	8.93	15.88	19.22	28.49
+65	39.74	43.89	4.36	5.85		
2022						
20-44	11.69	14.98	6.92	9.3		
45-64	38.53	56.84	11.5	18.8	24.07	36.07
+65	51.48	59.73	5.65	7.97		
2023						
20-44	14.99	13.10	8.87	8.13		
45-64	52.52	44.45	15.67	14.7	30.12	31.42
+65	50.85	64.43	5.58	8.59		

* Specific rate is computed multiplying crude rate by the proportion of the standard population, for women: 20-44 is 0.5919, 45-64 is 0.2984, +65 is 0.1097; for men: 20-44 is 0.6209, 45-64 is 0.33.07, +65 is 0.1334.

4. Discussion

After the cleaning step and outliers analysis, the total of records is 36,626,380 (99.89%), considering that the percentage of outliers records is 0.11% through the years, these were deleted. In the study case of this work, even a standardization rate was not applied, which seems to not significantly affect the AHRDM, but since the data scientist cannot assume that and just not to do the standardization, the opinion of the area where data science is applied must be considered in this work, epidemiology, because with another study population could not be the same, depends on the geographical area and people characteristics (like food habits, public policies about quality of care prevention of diabetes, etc.)

Figure 4 shows the specific crude rates versus the specific adjusted rates. Solid lines are for men and dashed lines are for women, for the 20-44 age group, in both cases, men and women, lower for the adjusted rates; for the 45-64 age group, also for men and women, the adjusted rates are lower, but the distance between rates is bigger, up to 50 AHRDM in some years; and for 65 and older, for both cases, adjusted rates are lower, but the distances are in some cases more than 100 AHRDM. Those results mean that standardization is required for specific rates. In Figure 5 the differences between the crude rates and the adjusted rates are shown, these were calculated with the highest rate minus the other rate.

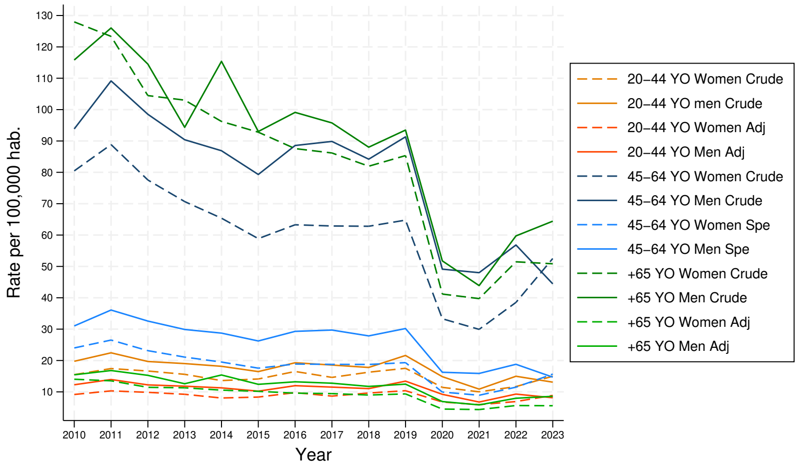

Figure 4. Rates by age and sex before and after the adjustment.

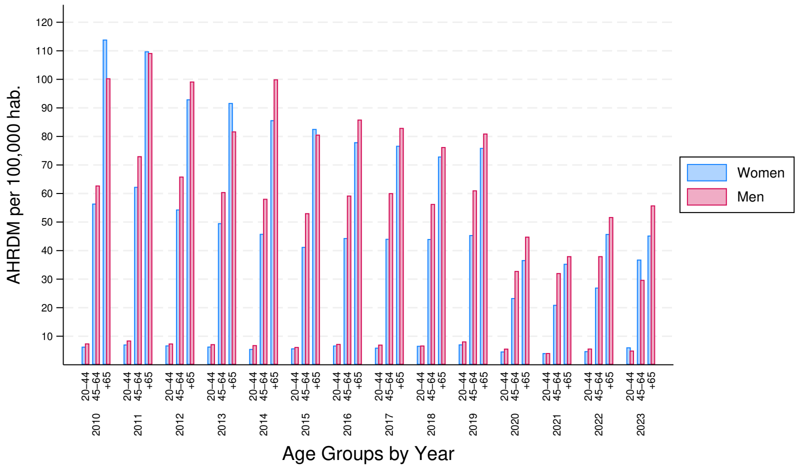

Figure 5. Differences between crude and adjusted rates.

5. Conclusions

This paper has shown the reasons why for a data scientist it is mandatory to learn about the field in which data science is applied and follow all the methods, processes, techniques, and transformations required in the field of knowledge to solve the problem. Even if as a data scientist you think you have enough knowledge to interpret the results, data science is evolving rapidly, but needs interdisciplinary scientists to apply it correctly.

The findings of this work prevent a common mistake when rates are presented without standardizing data. Non-standardized data are not valid when compared with other studies. Epidemiologists and specialized professionals, first of all, ask: data are standardized? As data scientists could answer yes because in the preparation data process, data could adjust computing the mean and standard deviation to calculate new values, but that is not the standardization to which epidemiologists are referring, that is said by experience, they mean to adjust by sex and age, at least, because other variables could be used.

Figure 5 also shows that, for this study, the standardization is notable for the 45-64 and 65 and older age groups, slightly similar for both groups, but for the 20-44 age group it is imperceptible, less than 10 AHRDM in all years. Taken together, these findings suggest that even the standardized data seem unnecessary, always must review all the methods of the field where data science is applied.

The most important limitation in this article lies in the fact that the data is only from Mexico. More research should focus on comparing methodologies for other diseases, even in another field not in health.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://www.dgis.salud.gob.mx/contenidos/basesdedatos/Datos_Abiertos_gobmx.html#. Also, an example of preprocessed data are available here: https://github.com/carloshn29/datasets

Acknowledgments: Thanks to CONAHCyT for the phd studies grant received during this research. Thanks to Instituto Politécnico Nacional and postgraduate and research section of the Interdisciplinary Professional Unit in Engineering And Advanced Technologies, UPIITA-IPN

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ICD-10-CM	International Classification for Diseases, 10th revision, Clinical Modification
AHRDM	Avoidable Hospitalizations Related to Diabetes Mellitus
DGIS	Dirección General de Informacion de Salud (due to terms of use, acronym must be in spanish
AHRQ	Agency for Healthcare Research and Quality
ACSC	Ambulatory Care-Sensitive Condition
IDE	Integrated Development Environment
PQI	Prevention Quality Indicator
MA	Metropolitan Area of Mexico City
SS	Secretaría de Salud (due to terms of use, acronym must be in spanish)

Appendix A

Table A1 shows in its columns, first, the year and age groups; second, number of MA cases and percentages by men and women; third, the MA populations of men and women for each age group.

Table A1. MA cases and populations by age groups.

Year	Metropolitan Area Cases		Population (habs.)	
Years-old	Women	Men	Women	Men
2010	3,611(100%)	3,460(100%)	8, 010,757	7,149,639
20-44	768(21.3%)	903(26.1%)	4,958,925	4,569,823
45-64	1,800(49.8%)	1,841(53.2%)	2,236,828	1,961,667
+65	1,043(28.9%)	716(20.7%)	815,004	618,149
2011	4,001(100%)	4,088(100%)	8,173,368	7,299,250
20-44	869(21.7%)	1,034(25.3%)	4,983,308	4,602,965
45-64	2,069(51.7%)	2,227(54.5%)	2,328,376	2,040,086
+65	1,063(26.6%)	827(20.2%)	861,684	656,199
2012	3,360(100%)	3,795(100%)	8,335,979	7,448,860
20-44	834(22.8%)	913(24.1%)	5,007,691	4,636,107
45-64	1,877(51.3%)	2,087(55.0%)	2,419,923	2,118,504
+65	949(25.9%)	795(20.9%)	908,365	694,249
2013	3,544(100%)	3,565(100%)	8,498,589	7,598,472
20-44	785(22.2%)	888(24.9%)	5,032,073	4,669,249
45-64	1,775(50.1%)	1,986(55.7%)	2,511,471	2,196,923
+65	984(27.8%)	691(19.4%)	955,045	732,300
2014	3,354(100%)	3,720(100%)	8,661,201	7,748,083
20-44	688(20.5%)	854(23.0%)	5,056,456	4,702,391
45-64	1,702(50.7%)	1,977(53.1%)	2,603,019	2,275,342
+65	964(28.7%)	889(23.9%)	1,001,726	770,350
2015	3,276(100%)	3,399(100%)	8,823,812	7,897,695
20-44	717(21.9%)	780(22.9%)	5,080,839	4,735,534
45-64	1,586(48.4%)	1,867(54.9%)	2,694,567	2,353,761
+65	973(29.7%)	752(22.1%)	1,048,406	808,400
2016	3,565(100%)	3,913(100%)	8,986,422	8,047,305
20-44	843(23.6%)	920(23.5%)	5,105,222	4,768,676
45-64	1,763(49.5%)	2,154(55.0%)	2,786,114	2,432,179
+65	959(26.9%)	839(21.4%)	1,095,086	846,450
2017	3,544(100%)	3,994(100%)	9,149,034	8,196,916
20-44	750(21.2%)	891(22.3%)	5,129,605	4,801,818
45-64	1,810(51.1%)	2,256(56.5%)	2,877,662	2,510,598
+65	984(27.8%)	847(21.2%)	1,141,767	884,500
2018	3,678(100%)	3,852(100%)	9,311,644	8,346,528
20-44	839(22.8%)	860(22.3%)	5,153,987	4,834,960
45-64	1,865(50.7%)	2,180(56.6%)	2,969,210	2,589,017
+65	974(26.5%)	812(21.1%)	1,188,447	922,551
2019	3,943(100%)	4,386(100%)	9,474,255	8,496,138
20-44	907(23.0%)	1,052(24.0%)	5,178,370	4,868,102
45-64	1,982(50.3%)	2,436(55.5%)	3,060,757	2,667,435
+65	1,054(26.7%)	898(20.5%)	1,235,128	960,601
2020	2,173(100%)	2,594(100%)	9,636,866	8,645,749
20-44	595(27.4%)	728(28.1%)	5,202,753	4,901,244
45-64	1050(48.3%)	1,349(52.0%)	3,152,305	2,745,854
+65	528(24.3%)	517(19.9%)	1,281,808	998,651
2021	2,023(100%)	2,348(100%)	9,799,477	8,795,360
20-44	524(25.9%)	537(22.9%)	5,227,136	4,934,386
45-64	971(48.0%)	1,356(57.8%)	3,243,853	2,824,273
+65	528(26.1%)	455(19.4%)	1,328,488	1,036,701
2022	2,607(100%)	3,036(100%)	9,962,088	8,944,970
20-44	614(23.6%)	744(24.5%)	5,251,519	4,967,528
45-64	1,285(49.3%)	1,650(54.3%)	3,335,400	2,902,691
+65	708(27.2%)	642(21.1%)	1,375,169	1,074,751
2023	3,314(100%)	2,697(100%)	10,124,698	9,094,582
20-44	791(23.9%)	655(24.3%)	5,275,901	5,000,670
45-64	1,800(54.3%)	1,325(49.1%)	3,426,948	2,981,110
+65	723(21.8%)	717(26.6%)	1,421,849	1,112,802

References

1. INEGI (Mexico), Press Release No. 657/22, november 10th **2022**.
2. Salas-Zapata L.; Palacio-Mejía L. S.; Aracena-Genao B.; Hernandez-Avila J. E.; Nieto-Lopez E. S., Costos Directos de las hospitalizaciones por diabetes mellitus en el Instituto Mexicano del Seguro Social. *Gaceta Sanitaria Vol. 32*, **2018**, *No. 3, Article 43*, 209-215 pages. ISSN 0213-9111.
3. Agudelo, M.; Murillo, J.; Gutierrez, L.; Giraldo, L.;*Hospitalizaciones y muertes evitables por condiciones sensibles a atención primaria en salud. México, 2005-2014*; Mexico, **2017**.

4. Flores, S.; Acosta, O.; Hernández, M.I.;Delgado, S.; Reyes, H.; Calidad de la atención en diabetes tipo 2, avances y retos de 2012 a 2018 2019 para el sistema de salud de México. *Salud Publica de Mexico*, **2020**. https://doi.org/10.21149/11876
5. Noncommunicable diseases. Available online: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed on 18 April 2024).
6. International Diabetes Federation. *IDF Diabetes Atlas*, 10th ed.; **2021**; ISBN: 978-2-930229-98-0.
7. Agency for Healthcare Research and Quality. Prevention Quality Indicator 93 (PQI 93), Prevention Quality Diabetes Composite. *AHRQ Quality indicators ICD-10-CM/PCS Specification* **2023**, U.S.
8. Goode, W.; Punjabi, V.; Niewiara, J.; Roberts, L.; Bruce, J.; Silva, S.; Morgan, B.; Pereira, K.; Brysiewicz, P.; Clarke, D.; *Using a Retrospective Secondary Data Analysis to Identify Risk Factors for Pulmonary Complications in Trauma Patients in Pietermaritzburg, South Africa*,, Journal of Surgical Research, Volume 262, **2021**, Pages 47-56, ISSN 0022-4804, U.S.; https://doi.org/10.1016/j.jss.2020.12.034.
9. Rattanavipapong, W.; Wang, Y.; Butchon, R.; Kittiratchakool, N.; Thammatacharee, J.; Teerawattananon, Y.; Isaranuwatchai, W., *Retrospective secondary data analysis to identify high-cost users in inpatient department of hospitals in Thailand, a middle-income country with universal healthcare coverage*, BMJ Open, **2021**, U.S.; https://doi:10.1136/bmjopen-2020-047330
10. Longbing Cao, Data Science: A Comprehensive Overview. *ACM Comput. Surv. Vol. 50*, **2017**, *No. 3, Article 43*, 42 pages.
11. Noncommunicable diseases. Available online: https://ec.europa.eu/eurostat/web/interactive-publications/demography-2023 (accessed on 18 April 2024).
12. Naing, N.N., Easy way to learn standardization: direct and indirect methods. *The Malaysian journal of medical sciences*, **2000**, *7(1)*. PMID: 22844209; PMCID: PMC3406211.
13. Higham, J.; Flowers, J.; Hall, P., Standardization. *Information on Public Health observatory recommended methods*, **2005**, *6*. ISSN: 1477-7290.
14. Merchant, A.T., Standardization. In: *Mitra, A.K. (eds) Statistical Approaches for Epidemiology*, 3rd ed.; Springer, Cham; **2024**; pp. 147–154; https://doi.org/10.1007/978-3-031-41784-9_9
15. Wood, S. M.; Yue, M., Kotsis, S. V.; Seyferth, A. V.; Wang, L.: Chung, K. C., Preventable Hospitalization Trends Before and After the Affordable Care Act. *AJPM Focus*, **2022**, *1(2), 100027*. https://doi.org/10.1016/j.focus.2022.100027
16. Saxena, A.; Ramamoorthy, V.; Rubens, M.; McGranaghan, P.; Veledar, E.; Nasir, K. Trends in quality of primary care in the United States, 2007-2016. *Scientific reports*, **2022**, *12(1), 1982*. https://doi.org/10.1038/s41598-022-06077-y
17. Keiding, N.; Clayton, D. Standardization and Control for Confounding in observational studies: A Historical Perspective. *Statistical Science*, **2014**, *29(4), 529-558*. https://doi.org/10.1214/13-STS453

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.